AFFTC-TIM-93-01

REVISION AND VERIFICATION OF A SEVEN-POINT WORKLOAD ESTIMATE SCALE

LAWRENCE L. AMES
Human Factors Engineer

EDWARD J. GEORGE
Human Factors Engineer

JULY 1993

TECHNICAL INFORMATION MANUAL

AIR FORCE FLIGHT TEST CENTER
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

AFFTC

This technical information memorandum (AFFTC-TIM-93-01) was developed as a non-funded research and development effort at the Air Force Flight Test Center (AFFTC), Edwards Air Force Base, California.

Prepared by:

LAWRENCE L. AMES
Human Factors Engineer

EDWARD J. GEORGE
Human Factors Engineer

This report has been reviewed and is approved for publication:

DEBORAH A. MUMMAW
Chief, Human Factors Branch

ROBERT W. RYAN, Lt Col, USAF
Chief, Engineering Division

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1993	Manual - Jan 1992 - Jun 1992

4. TITLE AND SUBTITLE
Revision and Verification of a Seven-Point Workload Estimate Scale

5. FUNDING NUMBERS
N/A

6. AUTHOR(S)
Lawrence L. Ames
Edward J. George

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
412 TW/DOEH
195 East Popson Ave
Edwards Air Force Base CA 93524-6841

8. PERFORMING ORGANIZATION REPORT NUMBER
AFFTC-TIM-93-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This technical information memorandum presents the results of an AFFTC effort to revise and verify the technical characteristics of a seven-point workload estimate scale. The original scale was The School of Aerospace Medicine (SAM) Form 202. From the results of the revision effort, it was concluded that the AFFTC revised workload estimate scale would be suitable for flight test applications.

14. SUBJECT TERMS
Subjective Workload Estimate
Subjective Workload
Workload Rating Scale

15. NUMBER OF PAGES
33

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

This page intentionally left blank.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

There has been a continuing need at the AFFTC for a simple, easy-to-use workload scale. Flight testing frequently required workload assessments from aircrew members and maintenance personnel. Test approaches and test plans often had to be developed quickly, not permitting scale development efforts during the test planning process. Aircrew ratings were sometimes required during flight, immediately following accomplishment of specific mission operations or test points. Absolute standards (i.e., pass-fail evaluation criteria) were sometimes specified as part of the test objectives, requiring absolute, rather than relative, workload assessment scales. The School of Aerospace Medicine (SAM) Form 202 contained a seven-point workload estimate scale that had many apparent advantages except that the scale suffered a serious drawback in that its psychometric characteristics had never been properly verified. This technical information memorandum (TIM) presents the results of an AFFTC effort to revise and verify the technical characteristics of a seven-point workload scale. From the results of the revision effort, it was concluded that the AFFTC revised workload estimate scale would be suitable for flight test applications in situations where an absolute assessment rather than a relative assessment of workload is desired, where an easy to understand scale is needed, where a minimum amount of subject training time is available, and where the collected data may be analyzed using statistical procedures requiring "interval" quality data.

1.0 INTRODUCTION

a. BACKGROUND

This technical information memorandum (TIM) presents the results of an AFFTC effort to revise and verify the technical characteristics of a seven-point workload scale. Data were collected from January 1992 through June 1992 from a total of 82 AFFTC test subjects. Data were collected by means of questionnaires and personal interviews.

There has been a continuing need at the AFFTC for a simple, easy-to-use workload scale. Flight testing frequently required workload assessments from aircrew members and maintenance personnel. Test approaches and test plans often had to be developed quickly, not permitting scale development efforts during the test planning process. Aircrew ratings were sometimes required inflight, immediately following accomplishment of specific mission operations or test points. Absolute standards (i.e., pass-fail evaluation criteria) were sometimes specified as part of the test objectives, requiring absolute, rather than relative, workload assessment scales.

The School of Aerospace Medicine (SAM) Form 202 (Appendix A) contained a seven-point workload estimate scale that had many apparent advantages except that the scale suffered a serious drawback in that its psychometric characteristics had never been properly verified. Advantages of the scale were that it was simple to use, required a minimum amount of pretest efforts, and the scale steps were anchored in absolute terms. On the other hand, the lack of verification of the technical characteristics meant that there was no assurance that the scale reflected a continuous underlying psychological dimension, that increasing scale steps reflected increasing levels of workload, or that the psychological intervals between scale steps were equal.

b. OBJECTIVE

The objective was to improve upon the original SAM Form 202 workload estimate scale and to verify the revised scale in terms of its ordinal and interval characteristics using pilots and other members of the flight test community as test subjects.

2.0 REVISION-VERIFICATION PROCESS

The scale development effort was performed iteratively and incrementally. The effort started with the original workload estimate scale (Appendix A). A definition of subjective workload was developed (Appendix B) to guide the scale revision and verification effort. Testing was accomplished through several cycles of assessing the psychological characteristics of the scale steps, revising the scale step definitions, and testing the magnitude of workload identified by these scale step definitions using psychometric methods. Pilots, engineers, weapons systems officers (WSOs), sensor operators, gunners, loadmasters, and maintenance personnel were used as test subjects, to reflect the intended subject population for flight testing. All pilots involved in testing were graduates of the USAF Test Pilot School, and most were on active duty.

a. REVISION

A starting point for scale development was the original SAM workload estimate (Table 1). Application of this scale has been reported (Reference 1). Prior study of the scale technical characteristics had been performed by George and Hollis. This unpublished study effort had shown some confusability between the scale step descriptors at the high workload end of the scale. From additional analysis performed by the present authors, four components of subjective workload were identified: activity level, system demands, time loads, and safety concerns. These components were incorporated into a definition of subjective workload (Appendix B). This definition served both to structure the scale development effort and to be used for subject training.

Scale revision was performed using guidance provided by Babbitt and Nystrom (Reference 2). The approach of having each scale step descriptor contain from two to four dimensions was retained from the original scale. Individual scale dimensions were refined to describe increasing workload magnitude. Scale step descriptor wording was revised in an attempt to produce subjectively equal intervals between steps and to reduce confusability between steps. Successive revisions were evaluated to bring the scale characteristics closer to an ideal straight line function. A straight line function would mean that the scale had both ordinal and interval characteristics. The three intermediate revisions used pair comparison testing and involved 33 test subjects, and was achieved through comments and written inputs from diverse sources including the developers of the original USAF SAM scale. The scale descriptors were then frozen for verification (Table 2). The definition of subjective workload (Appendix B) was considered as an integral part of the revised scale. The AFFTC revised workload estimate was then subjected to final verification testing. Two different psychometric methods were used for this verification testing, a pair compairson test and a rank order estimation test.

TABLE 1

THE USAF SAM WORKLOAD ESTIMATE SCALE

1) Nothing to do; No system demands.
2) Little to do; Minimum system demands.
3) Active involvement required, but easy to keep up.
4) Challenging, but manageable.
5) Extremely busy; Barely able to keep up.
6) Too much to do; Postponing some tasks.
7) Unmanageable; Potentially Dangerous; Unacceptable.

TABLE 2

THE AFFTC REVISED WORKLOAD ESTIMATE SCALE

1) Nothing to do; No system demands.
2) Light activity; Minimum demands.
3) Moderate activity; Easily managed; Considerable spare time.
4) Busy; Challenging but manageable; Adequate time available.
5) Very busy; Demanding to manage; Barely enough time.
6) Extremely busy; Very difficult; Non-essential tasks postponed.
7) Overloaded; System unmanageable; Essential tasks undone; Unsafe.

b. PAIR COMPARISON TEST

The method of pair comparisons has been described as a classic scale development method (Reference 3). However, the pair comparison procedure used was developed from the Subjective Workload Dominance (SWORD) procedure of Vidulich (References 4 and 5). Both the SWORD and the present pair comparison test approaches were based upon Saaty's broadly applicable analytic hierarchy process (Reference 6). For this effort, the SWORD procedure was modified to be used as a psychometric method. The SWORD procedure provided for a pairwise comparison of test items, so that for "N" items there would be N(N-1)/2 pairwise comparisons. Each pair comparison included an assessment of the degree of workload dominance of one item over the other. Typically, pair ratings could go from equal to maximum dominance (eight steps away on the worksheet). The pair comparison procedure used for this test included an analytic procedure developed by Turner (Reference 7) and a revised questionnaire incorporating a dominance scale adapted from Babbitt and Nystrom (Reference 2).

Pair comparison testing was conducted by self-administered questionnaire. Three alternative questionnaire forms were used to reduce order effects. Test subjects were selected on a quasi-random basis from among AFFTC flight test personnel, pilots, maintenance personnel and flight test engineers. Test subjects were given a questionnaire package (Appendix C) containing general instructions, the definition of subjective workload, and the pair comparison questionnaire. Subjects were instructed to read the workload definition and review the workload descriptors in the questionnaire prior to performing the ratings. The rating process (21 pair comparison ratings) took less than 30 minutes per subject.

c. RANK ORDER ESTIMATION TEST

Rank order techniques for scale verification have also been described as classic methods for scale development (Reference 3). Guilford described high positive correlations between the results of pair comparison and rank order procedures, indicating that they should produce similar results.

A one page questionnaire was developed (Appendix D) for the rank order estimation test. This test required separate judgments concerning rank order and relative interval distance of the scale descriptors. Twenty test subjects were selected from AFFTC flight test personnel, and five pilots, six engineers, gunners, and loadmasters were included. The test was administered through personal interview. Each of the seven scale step descriptors was printed on a separate four by five inch flash card, and each card was identified in one corner by a single letter of the alphabet.

During testing, test subjects were first required to rank order (sort) the workload descriptions on the flash cards from lowest to highest. Subjects then recorded this ranking on the questionnaire form, using the letter identifiers on the cards. Subjects were then asked to identify adjacent descriptors or terms that were confusable, and to describe any causes of confusion. This information was also recorded on the questionnaire form. Finally, subjects estimated the relative psychological distance (interval) between the scale steps by placing a check mark on a ruled line on the questionnaire form, ranging from 0 to 100. The highest and lowest workload levels were pre-defined at 0 and 100 respectively. On average, the rating process took about 5 minutes per subject.

This page intentionally left blank.

3.0 RESULTS

a. ORIGINAL SAM WORKLOAD ESTIMATE

The original SAM workload estimate scale was found to reflect a continuous underlying dimension, where the scale steps were ordinal but the psychological intervals between steps were unequal. Figure 1 shows the SAM workload estimate scale characteristics as determined from five test subjects using the pair comparison test approach. It should be noted that results from an ideal scale should plot as a perfectly straight line between zero (least workload) and one (most workload), demonstrating both perfect order and equal intervals between steps.

b. AFFTC REVISED WORKLOAD ESTIMATE

Figure 2 shows the results of the AFFTC revised workload estimate as determined by the two test procedures. The lower curve shows the mean results of the pair comparison test, while the upper curve shows the mean results of the rank order estimation test. Table 3 shows the ideal scale values assuming perfect linearity, and mean deviations of the obtained results from this ideal. This table shows that the combined results were closer to an ideal straight line function than the results of either test alone. Consequently, the combined results, shown in Figure 3, were used as the best estimate of the AFFTC revised workload estimate. Linear regression analysis was performed on the combined data (49 test subjects), and the following results were obtained. The correlation coefficient was 0.98, r squared was 96.4, and the standard error of estimate was 0.066. An analysis of variance by rating scale step produced an F ratio which was significant at less than the 0.0001 probability level. Analysis of the data from the 49 test subjects indicated a strong agreement between test subject ratings (Kendall's Coefficient of Concordance [w] was 0.997) (Reference 8). The combined test data from all 49 test subjects had a mean deviation from an ideal straight line of -1.16 percent, indicating that the obtained data were very close to ideal. Results from the test pilots using both test procedures were even closer to the ideal straight line function, having an average deviation from the ideal of 0.89 percent. The detailed test subject data and the results of the regression and analysis of variance test are presented in Appendix E.

Ordinal ranking of the scale steps was examined. Results for each test subject were plotted and compared with the group average. Only 3 of the 49 test subjects deviated from the ordinal ranking of the group, showing a 93.9% agreement between subjects. One subject exhibited an inversion between steps one and two, while two other subjects had zero differences in the rank value between steps rather than an expected step increase.

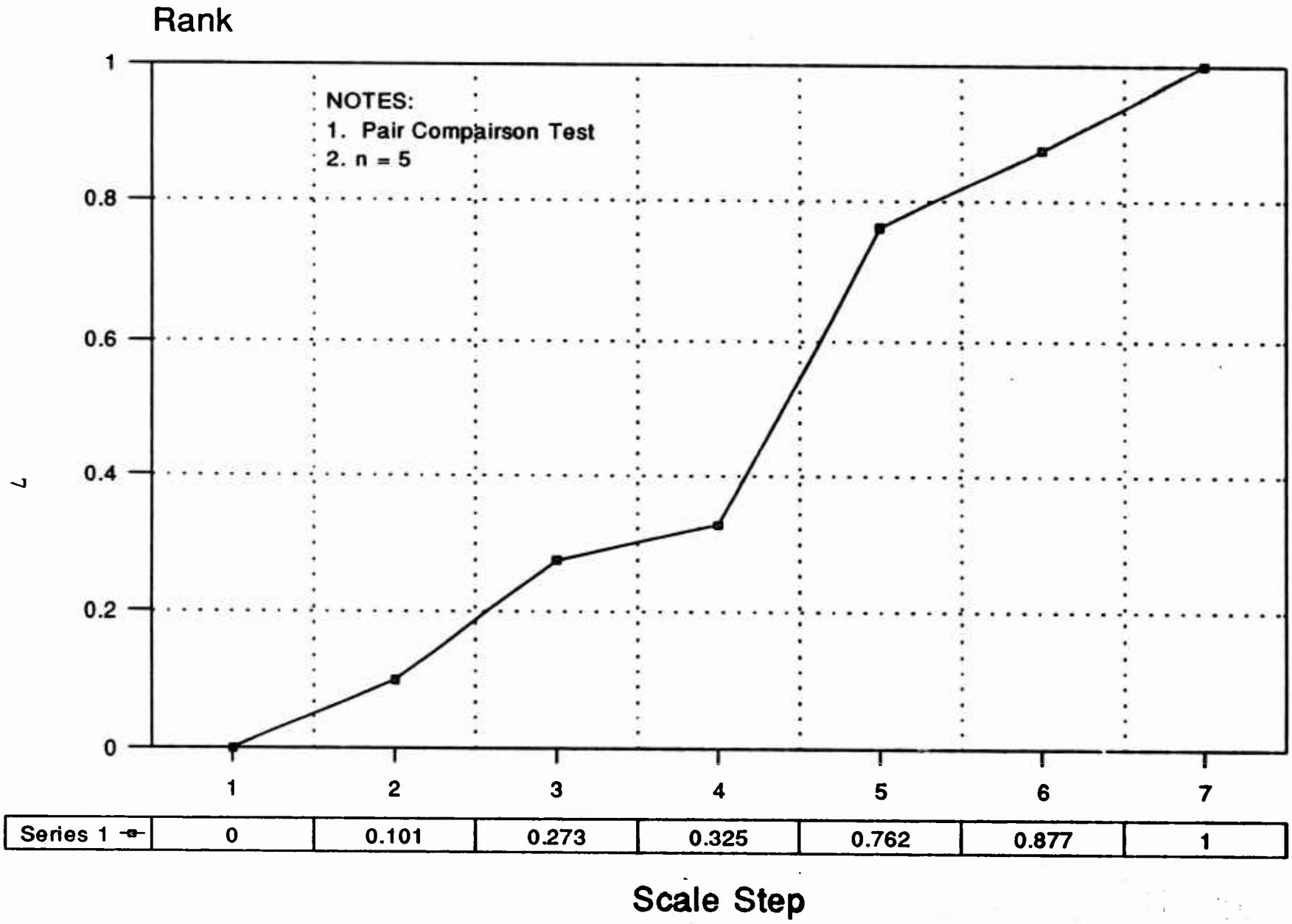
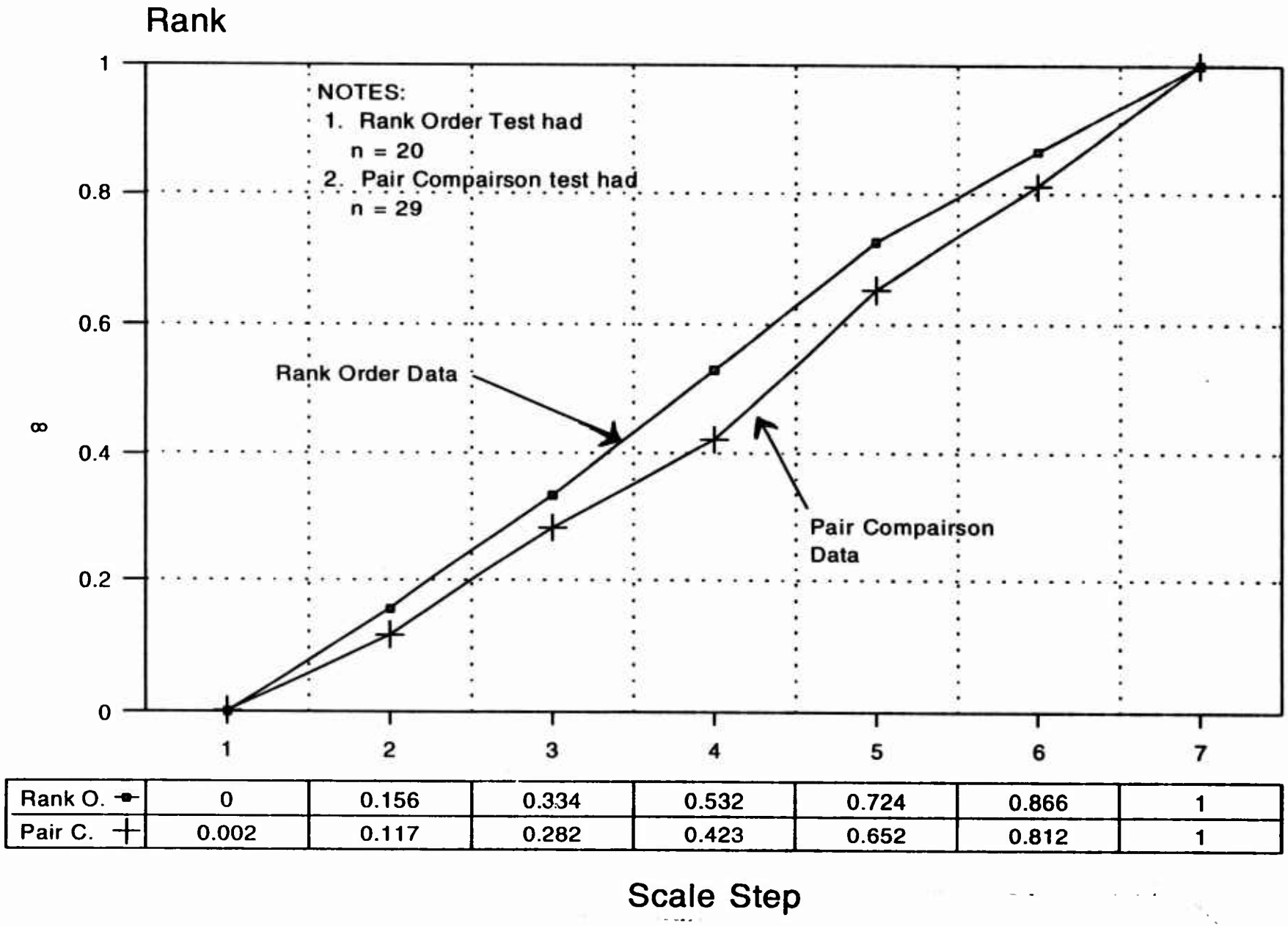


Figure 1 The Orginal USAF SAM Workload Estimate



Rank O.	0	0.156	0.334	0.532	0.724	0.866	1
Pair C.	0.002	0.117	0.282	0.423	0.652	0.812	1

Figure 2 The AFFTC Revised Workload Estimate Assessed by Two Methods

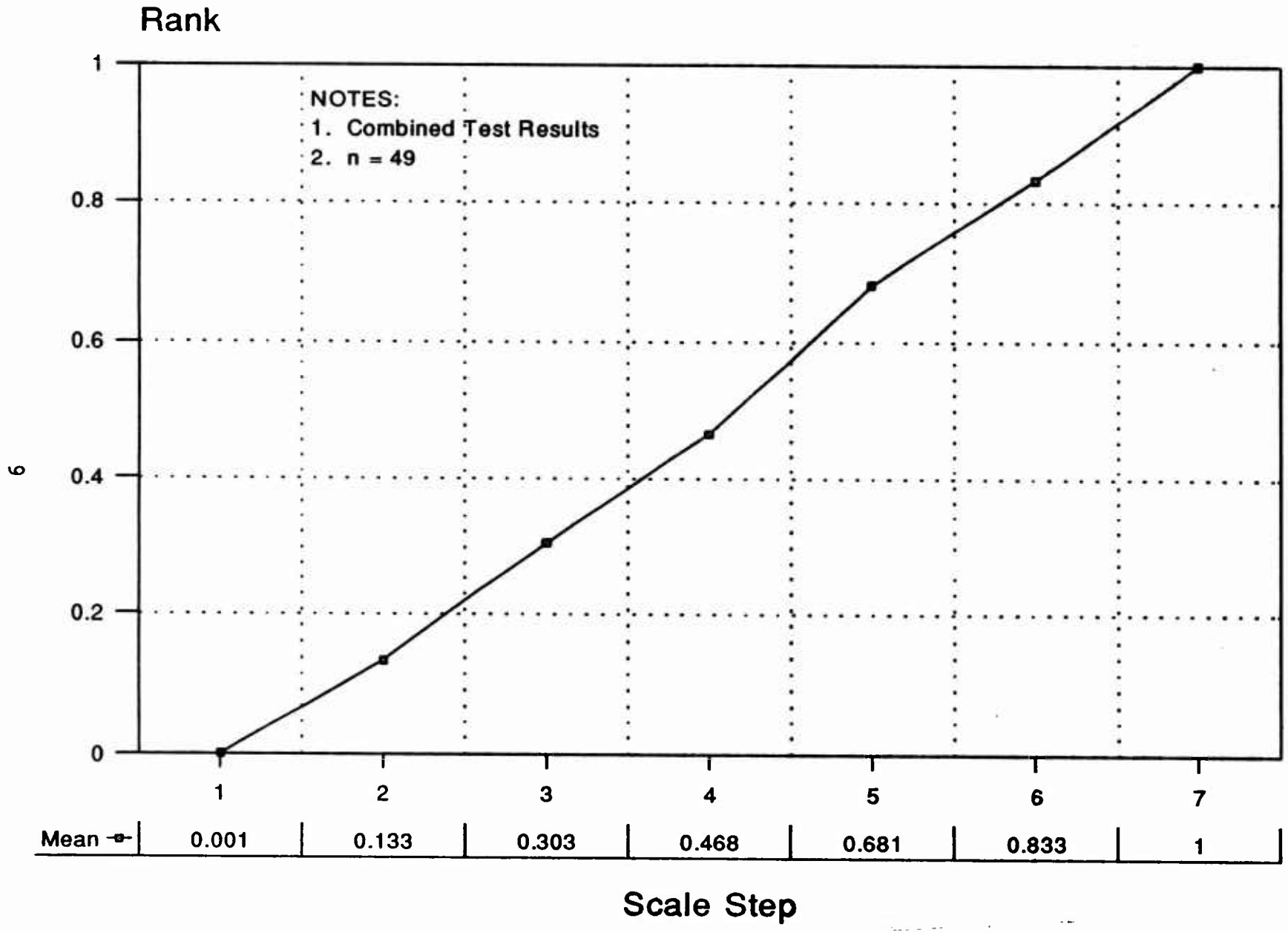


Figure 3 The Final AFFTC Revised Workload Estimate

Table 3

MEAN TEST RESULTS OF THE AFFTC REVISED WORKLOAD ESTIMATE

Scale Step	Ideal Value	Pair Comparison Results	Rank Order Results	Combined Results
1	0.000	0.002	0.000	0.001
2	0.166	0.117	0.156	0.133
3	0.333	0.282	0.334	0.303
4	0.500	0.423	0.532	0.468
5	0.666	0.652	0.724	0.681
6	0.833	0.812	0.866	0.833
7	1.000	1.000	1.000	1.000
Sample Size	=	29	20	49
Deviation from Ideal (Mean %)	=	-3.03	1.62	-1.16

Confusability data was collected only from the rank order test subjects. Eight instances of confusability between workload step descriptors were reported by the 20 test subjects during testing. Of these instances of confusability, six involved response alternatives "one" versus "two". The remaining two instances involved response alternatives "four" versus "five", and "five" versus "six". All eight instances were the result of slight similarities between sub-dimensions of the compared response alternative definitions rather than any real confusion between the overall definitions themselves. Thus, these instances of confusion were not strong enough to seriously compromise the subjective distance between adjacent response alternatives or to cause the rank ordering between response alternatives to be altered. It was speculated that the above mentioned confusability caused the interval separation between response alternatives to deviate slightly from a perfectly linear function. Given the difficulties inherent in attempting to express precise magnitude relations with non-numerical terms, contamination of the observed kind may be impossible to completely eliminate from this subjective scale.

This page intentionally left blank.

4.0 DISCUSSION

A theoretical concept behind this effort was that workload was defined as a multi-dimensional concept, where individual raters could implicitly integrate the various workload dimensions into a single value along some unidimensional continuum. When assessing a task, the rater could use only one workload dimension or may mentally combine the psychological contributions of two, three, or all four dimensions to arrive at a single number (from one to seven) describing their subjective experience of workload. This theoretical concept raised several issues about the scale itself as well as being a topic for additional study.

One issue was whether workload should be considered a uni-dimensional or multi-dimensional concept. Moray provides convincing support for a multi-dimensional concept of workload (Reference 9). From this paper, it was not clear exactly what the dimensions should be or how the assessments should be combined to describe workload. The multi-dimensional approach of the original USAF SAM workload estimate scale was retained for the revision effort because the objective of the effort was to emphasize evolutionary development rather than radical change. The present authors believe that the individual raters were able to integrate the various workload factors into a single (i.e., unidimensional) workload rating ranging from 1 (least workload) to 7 (most workload).

Reliability and validity were topics that should be studied for the revised workload estimate. The present effort was concerned with the scale descriptors themselves. Inter-rater reliability using the scale descriptors was found to be quite high (Kendall's Coefficient of Concordance [w] was 0.997). However, reliability assessments should be obtained using the revised workload scale to assess workload using ratings of actual job task performance. Also, validity studies should be made comparing the results of the revised workload estimate with the results of other, more proven, workload assessment tools.

The present effort used two different verification procedures as a form of cross-check. In the pair comparisons test, the raters were not told how many scale steps there were, made 21 comparative ratings on unnumbered scales, and had no feedback of their results during the test. Scale ordinality and the intervals between scale steps were determined mathematically from the subject ratings, and the test subjects had no direct feedback of their performance. In the rank order test, the subjects knew exactly how many scale steps were being evaluated, that the descriptors were intended to be ordered on a continuum, and that two descriptors that they selected were to be anchored at the scale ends. The rank order procedure subjects had immediate visual feedback as to their rank order judgments and the relative spacings they had indicated between the scale steps, and could correct their responses if desired. However different the procedures were, the obtained results were remarkably similar, having a Pearson Product Moment Correlation (r) of +0.994. The similarity of the results of the two procedures supported use of the

combined data as the best estimate of the scale characteristics, and furthermore, supported the validity of the overall scale development effort.

The analytic results of the revised workload scale indicated nearly linear increases in rank value by scale step. This result indicated that the data resulting from use of the revised scale may be considered as "interval" quality. Data has been identified in ascending quality as; nominal, ordinal, interval, and ratio by S.S. Stevens (Reference 10). The issue of data quality and "permissible statistics" has been discussed from various theoretical perspectives by J. Mitchell (Reference 11).

5.0 CONCLUSION

The test objectives were met. The AFFTC revised workload estimate was found to be an improvement over the original SAM Form 202 workload estimate, and the revised scale was found to have the expected ordinal characteristics across scale steps with nearly equal psychological intervals between workload steps. The scale characteristics of the revised scale were nearly ideal, so that the data obtained from the use of this scale may be considered as interval quality. The fact that verification testing included pilots and other aircrew members supported the potential usefulness of this scale for flight test applications. It was concluded that the AFFTC revised workload estimate scale would be suitable for flight test applications in situations where an absolute assessment rather than a relative assessment of workload is desired, where an easy to understand scale is needed, where a minimum amount of subject training time is available, and where the collected data may be analyzed using statistical procedures requiring "interval" quality data. Additional studies should be performed in the future to assess the reliability and validity of the AFFTC revised workload estimate using the scale to assess workload within flight test applications.

REFERENCES

1. Gawron, V.J., Schiflett, S., Miller, J., Ball, J., Slater, T., Parker, F., Lloyd, M., Travele, D., and Spicuzza, R.J. *The Effect of Pyridostigmine Bromide on Inflight Aircrew Performance* (USAF-SAM-TR-87-24). Brooks Air Force Base, TX: School of Aerospace Medicine, January 1988.

2. Babbitt, B. A., Nystrom, C.O. *Questionnaire Construction Manual.* U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia: ARI Research Project 89-20, June 1989.

3. Guilford, J.P. *Psychometric Methods.* New York: Mc Graw-Hill Book Company, Inc., 1954.

4. Vidulich, M. A., Ward, G. F. , Schueren, J. Using the Subjective Workload Dominance (SWORD) Technique for Projective Workload Assessment. *Journal of the Human Factors Society, Vol. 33, No. 6* (pp 677-691), December 1991.

5. Vidulich, M.A. The use of Judgment Matrices in Subjective Workload Assessment. In: *Proceedings of the Human Factors Society 33rd Annual Meeting, Volume 2* (pp 1406-1410), Santa Monica, California: Human Factors Society.

6. Saaty, T.L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation.* New York, N.Y.: McGraw Hill, 1980.

7. *AFFTC Human Factors Engineering Training and Reference Handbook and Data.* U.S. Air Force Flight Test Center, Edwards Air Force Base, California, 22 Nov. 91.

8. Hayes, W.L. *Statistics for the Social Sciences.* New York: Holt, Rinehart and Winston, Inc., 1973.

9. Moray, N. Mental Workload Since 1979. In: *International Reviews of Ergonomics*, David J. Osborne (ed.) Volume 2. (pp 123-150) London: Taylor and Francis, 1988.

10. Stevens, S.S. Mathematics, measurement and psychophysics. In S.S. Stevens (ed.) *Handbook of Experimental Psychology* (pp. 1- 49). Wiley: New York, 1951.

11. Michell, J. *Measurement Scales and Statistics: A Clash of Paradigms.* Psychological Bulletin, Vol. 100, No. 3 (pp. 398-407), 1986.

APPENDIX A

CREW STATUS CHECK (SAM FORM 202)

NAME	DATE AND TIME

SUBJECTIVE FATIGUE

(Circle the number of the statement which describes how you feel RIGHT NOW.)

1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not At Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

COMMENTS

WORKLOAD ESTIMATE

(Circle the number of the statement which best describes the MAXIMUM workload you experienced during the PAST HOUR. Estimate and record the number of MINUTES during the past hour you spent at this workload level.)

		MINUTES
1	Nothing to do; No System Demands	
2	Little to do; Minimum System Demands	
3	Active Involvement Required, But Easy to Keep Up	
4	Challenging, But Manageable	
5	Extremely Busy; Barely Able to Keep Up	
6	Too Much to do; Overloaded; Postponing Some Tasks	
7	Unmanageable; Potentially Dangerous; Unacceptable	

COMMENTS

SAM FORM JUL 80 202 CREW STATUS CHECK

APPENDIX B

DEFINITION OF SUBJECTIVE WORKLOAD

SUBJECTIVE WORKLOAD

Subjective workload is a multi-dimensional concept. For the AFFTC revised workload estimate, a wide variety of contributing factors are identified within four areas. Subjective workload increases as the demands in any one or more of these areas increase. This scale approach requires you, the worker, to integrate the contributing factors to workload and arrive at an overall workload rating from least (1) to most (7).

ACTIVITY LEVEL: Activity level may range from nothing to do to an overwhelming amount to do. Worker actions may include locomotion, arm and leg movements, and manual manipulation. Physical activity becomes more complex as task action variety increases, and as the physical locus of action shifts from place to place. High levels of physical activity may act to stress muscles, deplete energy reserves, cause tiredness and fatigue, and eventually lead to total exhaustion.

SYSTEM DEMANDS: Task demands may range from simple and repetitive to complex and demanding. Difficult tasks may involve sensing things that are hard to see or difficult to hear, require extreme concentration to overcome distractions, involve detailed memory or thought, and require important decisions to be made. Tasks may also require precise hand-eye control or multi-limb coordination. In addition, the working environment may include conditions which make work difficult, such as; extremes of hot or cold, high humidity levels, distracting noise or vibration, and poor air quality. Physical conditions of the worker may also increase workload, such as lack of sleep or rest, inadequate food or water intake, or inadequate or unappealing workspace.

TIME LOADS: The amount of time available to accomplish tasks may vary from plentiful to non-existent. Inadequate time available for task completion stresses workers, increasing workload. When little time is available, multiple tasks may have to be prioritized mentally and acted upon with haste, often resulting in mistakes that require work to be re-done. Sometimes tasks may have to be postponed or even ignored completely. The resulting confusion and frustration further increase workload.

SAFETY CONCERNS: Concern for personal physical safety, or the responsibility of protecting equipment or supplies from damage, increases subjective workload. Safety concerns are high when situations are inherently dangerous and life-threatening. Other situations may be dangerous and stressful because the operator cannot see or hear needed information, or because the system design does not permit adequate control of knowledge of results of control actions.

APPENDIX C

PAIR COMPARISON QUESTIONNAIRE PACKAGE WITH ANALYSIS WORKSHEET

PAIR COMPARISONS QUESTIONNAIRE FORM

FOR REVISED WORKLOAD ESTIMATE

Consider each descriptor pair below. If equal, put a check in the left-most column. If unequal, circle the letter of the descriptor describing the **higher** level of workload and rate the degree of unequalness by checking one of the other eight columns.

DESCRIPTOR PAIR	RELATIVE WORKLOAD DOMINANCE EQUAL	A Little More	A Good Deal More	Very Much More
a. Very busy; Demanding to manage; Barely enough time. b. Busy; Challenging but manageable; Adequate time available.				
a. Busy; Challenging but manageable; Adequate time available. b. Extremely busy; Very difficult; Non-essential tasks postponed.				
a. Extremely busy; Very difficult; Non-essential tasks postponed. b. Light activity; Minimum demands.				
a. Busy; Challenging but manageable; Adequate time available. b. Nothing to do; No system demands.				
a. Very busy; Demanding to manage; Barely enough time. b. Overloaded; System unmanageable; Essential tasks undone; Unsafe.				
a. Light activity; Minimum demands. b. Very busy; Demanding to manage; Barely enough time.				

RELATIVE WORKLOAD DOMINANCE

DESCRIPTOR PAIR	EQUAL	A Little More	A Good Deal More	Very Much More
a. Nothing to do; No system demands.				
	—	— — — — — — — —		
b. Overloaded; System unmanageable; Essential tasks undone; Unsafe.				
a. Extremely busy; Very difficult; Non-essential tasks postponed.				
	—	— — — — — — — —		
b. Overloaded; System unmanageable; Essential tasks undone; Unsafe.				
a. Extremely busy; Very difficult; Non-essential tasks postponed.				
	—	— — — — — — — —		
b. Very busy; Demanding to manage; Barely enough time.				
a. Extremely busy; Very difficult; Non-essential tasks postponed.				
	—	— — — — — — — —		
b. Moderate activity; Easily managed; Considerable time to spare.				
a. Busy; Challenging but manageable; Adequate time available.				
	—	— — — — — — — —		
b. Overloaded; System unmanageable; Essential tasks undone; Unsafe.				
a. Busy; Challenging but manageable; Adequate time available.				
	—	— — — — — — — —		
b. Moderate activity; Easily managed; Considerable spare time.				
a. Moderate activity; Easily managed; Considerable spare time.				
	—	— — — — — — — —		
b. Very busy; Demanding to manage; Barely enough time.				

DESCRIPTOR PAIR	RELATIVE WORKLOAD DOMINANCE EQUAL	A Little More	A Good Deal More	Very Much More
a. Moderate activity; Easily managed; Considerable spare time.				
b. Light activity; Minimum demands.	— |	— — — — — — — —		
a. Overloaded; System unmanageable; Essential tasks undone; Unsafe.				
b. Moderate activity; Easily managed; Considerable spare time.	— |	— — — — — — — —		
a. Light activity; Minimum demands.				
b. Busy; Challenging but manageable; Adequate time available.	— |	— — — — — — — —		
a. Nothing to do; No system demands.				
b. Light activity; Minimum demands.	— |	— — — — — — — —		
a. Very busy; Demanding to manage; Barely enough time.				
b. Nothing to do; No system demands.	— |	— — — — — — — —		
a. Nothing to do; No system demands.				
b. Moderate activity; Easily managed; Considerable spare time.	— |	— — — — — — — —		
a. Nothing to do; No system demands.				
b. Extremely busy; Very difficult; Non-essential tasks postponed.	— |	— — — — — — — —		
a. Overloaded; System unmanageable; Essential tasks undone; Unsafe.				
b. Light activity; Minimum demands.	— |	— — — — — — — —		

Check mark to numerical data entry translation guide:

0	1	2	3	4	5	6	7	8
—	—	—	—	—	—	—	—	—

DATA SUMMARY SHEET

	1	2	3	4	5	6	7	SUM	RANK
1	0								
2		0							
3			0						
4				0					
5					0				
6						0			
7							0		

NOTE: Ratings only half-fill the matrix. Fill out the matrix with complementary numbers. Sum rows. Compute rank using formula at the right.

$$\text{Rank} = \frac{(X - \text{Min})}{(\text{Max} - \text{Min})}$$

APPENDIX D

RANK ORDER QUESTIONNAIRE EXAMPLE

WORKLOAD SCALE SURVEY

The following survey is intended to solicit information concerning your perceptions of a proposed workload scale. The survey is made up of several tasks. Please complete these tasks in the order identified below.

TASK ONE: Fill in your name and your primary job title in the spaces provided below:

Name:_____ Job Title: _____

TASK TWO: Find enclosed seven flash cards. Each flash card has a phrase written on it which defines a level of workload.

1. Taking careful note of the content of each definition, sort the cards on a line in front of you, placing the lowest level of workload on the far left and the highest level of workload on the far right.

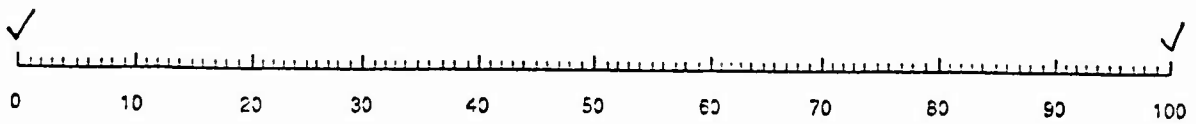
2. Each card has a letter of the alphabet affixed in the upper left hand corner. When you are satisfied that you have sorted the cards correctly, write their respective letters in the seven boxes provided below, one letter for each box, with the letter for the lowest level of workload in the first box, the next higher level of workload in the second box, and so on.

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH

TASK THREE: Identify any two adjacent scale definitions that you think are confusable with one another. In other words, if you think the forth definition defines the same or nearly the same level of workload as the fifth definition, then write below, "4 with 5" and so on. If none of the response alternatives are confusable then check the box labeled "NO".

NO ( ) ___ with ___, ___ with ___, ___ with ___.

TASK FOUR: Estimate the amount of workload defined by each of the seven scale definitions. Imagine that the lowest level of workload (first alternative) was placed on the line below at the point checked as "0", and the highest level of workload (seventh alternative) at the point checked as "100". With this in mind, check a point on the line between 0 and 100 for each of the other five response alternatives according to where you think they belong relative to one another.



APPENDIX E

TEST DATA BY SUBJECT AND TEST PROCEDURE

Table E1

MEAN RANKS FOR PAIR COMPARISON TEST

	Scale Step						
Subject Number	1	2	3	4	5	6	7
401	0.000	0.113	0.209	0.467	0.629	0.693	1.0
402	0.000	0.182	0.288	0.424	0.636	0.757	1.0
403	0.000	0.098	0.246	0.377	0.738	0.869	1.0
404	0.000	0.066	0.317	0.533	0.666	0.850	1.0
407	0.000	0.091	0.348	0.500	0.651	0.909	1.0
408	0.000	0.092	0.277	0.492	0.615	0.754	1.0
409	0.000	0.078	0.297	0.531	0.734	0.859	1.0
410	0.000	0.145	0.223	0.302	0.631	0.763	1.0
411	0.000	0.125	0.312	0.328	0.781	0.844	1.0
412	0.000	0.079	0.269	0.365	0.508	0.777	1.0
414	0.000	0.200	0.382	0.382	0.600	0.873	1.0
415	0.000	0.096	0.274	0.370	0.616	0.712	1.0
416	0.000	0.132	0.415	0.472	0.868	0.943	1.0
417	0.000	0.166	0.333	0.500	0.666	0.833	1.0
418	0.000	0.033	0.217	0.317	0.617	0.850	1.0
419	0.000	0.246	0.461	0.477	0.661	0.815	1.0
420	0.000	0.127	0.222	0.444	0.682	0.746	1.0
421	0.000	0.131	0.131	0.342	0.645	0.802	1.0
422	0.000	0.203	0.390	0.474	0.712	0.898	1.0
423	0.000	0.137	0.274	0.397	0.562	0.698	1.0
424	0.055	0.000	0.219	0.342	0.575	0.726	1.0
425	0.000	0.085	0.268	0.329	0.500	0.719	1.0
426	0.000	0.156	0.234	0.415	0.610	0.766	1.0
427	0.000	0.061	0.231	0.461	0.692	0.785	1.0
428	0.000	0.091	0.303	0.394	0.621	0.879	1.0
432	0.000	0.082	0.246	0.442	0.836	0.951	1.0
440	0.000	0.119	0.254	0.373	0.508	0.830	1.0
441	0.000	0.156	0.281	0.500	0.687	0.765	1.0
445	0.000	0.097	0.274	0.516	0.661	0.839	1.0
Mean =	0.002	0.117	0.282	0.423	0.652	0.812	1.0
s (1) =	0.010	0.052	0.069	0.070	0.086	0.072	0.0
C (2) =	50.0	44.4	24.5	16.5	13.2	8.9	0.0

NOTES:
1. "s" means Standard Deviation.
2. "C" means Coefficient of Variation.
3. Subjects 426, 427, and 428 were pilots.
4. Subject numbers start at 400 to represent the fourth test iteration.

Table E2

MEAN SCALE VALUES FOR RANK ORDER TEST

	Scale Step						
Subject Number	1	2	3	4	5	6	7
S1	0.0	0.14	0.30	0.60	0.80	0.92	1.0
S2	0.0	0.17	0.36	0.50	0.67	0.86	1.0
S3	0.0	0.20	0.40	0.60	0.80	0.90	1.0
S4	0.0	0.10	0.20	0.50	0.75	0.90	1.0
S5	0.0	0.17	0.35	0.50	0.66	0.80	1.0
S6	0.0	0.15	0.50	0.60	0.75	0.90	1.0
S7	0.0	0.10	0.30	0.50	0.60	0.75	1.0
S8	0.0	0.20	0.40	0.60	0.80	0.90	1.0
S9	0.0	0.15	0.34	0.50	0.70	0.85	1.0
S10	0.0	0.30	0.35	0.60	0.80	0.90	1.0
S11	0.0	0.15	0.35	0.50	0.65	0.85	1.0
S12	0.0	0.10	0.25	0.50	0.70	0.90	1.0
S13	0.0	0.10	0.30	0.50	0.75	0.90	1.0
S14	0.0	0.10	0.20	0.30	0.60	0.80	1.0
S15	0.0	0.20	0.35	0 .50	0.70	0.90	1.0
S16	0.0	0.10	0.30	0.50	0.70	0.90	1.0
S17	0.0	0.20	0.40	0.60	0.85	0.90	1.0
S18	0.0	0.15	0.35	0.65	0.85	0.90	1.0
S19	0.0	0.20	0.38	0.60	0.70	0.80	1.0
S20	0.0	0.15	0.30	0.50	0.65	0.80	1.0
Mean =	0.0	0.156	0.334	0.532	0.724	0.866	1.0
s (1) =	0.0	0.051	0.070	0.076	0.075	0.050	0.0
C (2) =	N/A	32.7	21.0	14.3	10.4	5.8	0.0

NOTES:
1. "s" means Standard Deviation.
2. "C" means Coefficient of Variation.
3. Subjects S1, S8, S10, S19, and S20 were pilots.
4. Data are presented from 0 to 1, rather than from 0 to 100 as originally collected.

TABLE E3

RESULTS OF THE REGRESSION AND ANALYSIS OF VARIANCE TEST OF THE COMBINED DATA

Regression Analysis - Linear Model

Parameter	Estimate	Standard Error	Student's t Value	Probability Level
Intercept	-0.19358	8.01152E-3	-24.1627	0.00000
Slope	0.17054	1.79143E-3	95.1989	0.00000

Dependent Variables: Ratings

Independent Variable: Levels

Analysis of Variance - Seven Level (scale step) Model

Source	Sum of Squares	Df	Mean Square	F-Ratio	Probability Level
Model	39.9042	1	39.9042	9062.84	0.00000
Error	1.5014	341	0.0044		
Total	41.4056	342			

Correlation Coefficient = 0.981702

R-Squared = 96.37 Percent

Standard Error of Estimate = 0.0663555